DOCUMENT RESUME

ED 282 888 TM 870 239

AUTHOR Craig, James R.; And Others
TITLE Evaluating Effective Teaching in Colleges and Universities: How Far Have We Come?
PUB DATE 86
NOTE 23p.; Paper presented at the Annual Meeting of the American Evaluation Association (Kansas City, MO, October 29-November 1, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Classroom Techniques; Cognitive Processes; *College Faculty; *Evaluation Criteria; Evaluation Methods; *Evaluation Problems; Higher Education; Information Needs; Rating Scales; *Student Evaluation of Teacher Performance; *Teacher Effectiveness; *Teacher Evaluation

ABSTRACT

An examination of the policies and processes used to evaluate teaching in colleges and universities over the past several decades begins by identifying six key features of the research literature: (1) no formal definition of effective teaching has been generated to guide instrument development, data collection, and interpretation; (2) the purpose of the evaluation affects the form and substance of the evaluation process; (3) the criteria for the determination of validity have not been adequately defined; (4) direct observation of instruction is almost never done in higher education; (5) a large number of studies have been conducted to identify frame factors that must be considered when interpreting students' ratings of instructors; and (6) problems in interpreting student ratings are documented. The most common criterion for demonstrating validity of students' ratings is a measure of student achievement; however, its validity for predicting teacher effectiveness has not been tested. Different evaluation audiences--administrators, teachers, and students--have different information needs which affect data collection. Researchers are developing solutions for these problems. Their approaches include developing rating scales for multiple audiences, examining cognitive processes in the classroom and time on task, and studying students' attitudes and perceptions. A 40-item reference list is included. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED282088

Evaluating Effective Teaching in Colleges and Universities: How Far Have We Come?

James R. Craig, Doris L. Redfield, and Gary R. Galluzzo

Western Kentucky University

TM 870 239

Running head: EVALUATING EFFECTIVE TEACHING

Paper presented at the 1986 Annual Meetings of the American Evaluation Association, Kansas City.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. R. Craig

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

Evaluating Effective Teaching in Colleges and Universities: How Far Have We Come?

When examining the procedures and processes used to evaluate teaching in colleges and universities over the past several decades, one is simultaneously impressed with how far we have come and, at the same time, how little progress we have made. The following paragraphs provide documentation for those impressions.

Brief Overview of Where We Have Been

The research literature dealing with the evaluation of teaching in colleges and universities is quite large, is spread across a number of disciplines, and is published in a variety of journals, monographs, newsletters, books, and magazines. As a result, it is virtually impossible to be aware of all that has been written dealing with the topic. In attempting to synthesize the research literature in this area, we have identified six key features of the research literature on evaluating teaching in higher education.

One of the most striking features of the research on evaluation of teaching at colleges and universities is that no formal definition of effective teaching has been generated and is generally accepted to guide instrument development, data collection, and interpretation. Rather, definition of effective teaching has usually been accomplished indirectly through the specification of different educational outcomes. These outcomes in turn represent different sets of criteria for judging the instructional effectiveness of the teacher (McKeachie, 1979). Effective teaching, therefore, ends up being defined in a reverse

fashion via the data collection processes and procedures we developed. That is, the data collection instruments define what effective teaching is rather than our definitions of effective teaching specifing the form and substance of the data collection instruments.

A related problem is the nature of the purpose of the evaluation -- personal development (often tabbed formative evaluation), personnel management (often termed summative evaluation), or consumer (i.e., student) information. The purpose of the evaluation determines to a large degree the form and substance of the evaluation process (Braskamp, Brandburg, & Ory, 1984) and, therefore, what is termed to be effective teaching.

The third general feature is related to the assessment of the reliability and validity of data collection instruments used to evaluate teaching. In general, the reliability of the various forms has been found to be acceptable (e.g., Kinicki, Bannister, Hom, & Denisi, 1985). However, the concern for validity has not been clearly answered (McKeachie, 1979). The criteria for the determination of validity have not been adequately defined in part because there have not been clear, generally acceptable definitions of effective teaching and the components of it.

The fourth feature of the research on evaluation of teaching at colleges and universities is that, while there are exceptions (e.g., Murray, 1983), direct observation of instruction is almost never done in higher education. Rather, some type of observation-by-proxy is used -- such as student ratings (Braskamp, Brandenburg, & Ory, 1984). Whatever the data

capturing procedure, usually the attempt is made to keep the data as objective as possible so they are relatively easy to count.

A fifth feature is that a large number of studies have been conducted to identify frame factors -- factors that must be considered when interpreting students' ratings of instructors. The research may be summarized in terms of two primary outcomes. First, a number of variables have been identified that significantly influence students' ratings of teachers such as class size (e.g., Centra, 1983) or expected grade (e.g., Feldman, 1983). Second, studies of frame factor variables have identified things that instructors can do, which may or may not be related to more effective teaching, to improve their ratings such as teacher self-disclosure (McCarthy, & Schmeck, 1982).

Related to the identification of frame factors has been the documentation of problems of interpreation of student ratings -- the sixth identifiable feature of the research on evaluation of teaching in colleges and universities. For instance, it has been found that written student comments sometimes have greater influence on decision makers than statistical summaries of a number of students' ratings (e.g., Lin, McKeachie, & Tucker, 1984).

Where Are We Now?

So, given these impressions, where are we now? The answer probably is that we are on the road to problem definition. We have recognized that effective teaching and its assessment is a multidimensional problem (Marsh, 1984; Miller, 1984; Shields, 1984). But, in considering the problem, it is helpful to consider what the problem is not and it must be made very clear that the

problem is not that we cannot count. We can count many things. Scholarly activities and service events can be, and are, counted (with varying degrees of defensibility). The problem is not that we do not have the technical expertise to devise reliable student questionnaires. The evaluation literature abounds with instrument development and reliability studies. The problem is one of validity. "For what purposes are student responses to course evaluation questionnaires valid? At present, the most defensible response to the problem is probably, "It depends." -- It depends upon: (a) the criterion for validity and (b) the evaluation audience.

The Criterion for Validity

If the intent of student course evaluations is to provide a measure of teaching effectiveness, then the criterion or criteria for effectiveness must be identified and defined. Most validity studies have used a measure of student achievement as the criterion variable. With unconvincing exception, these studies have yielded moderate correlations between questionnaire responses and achievement (e.g., Centra, 1977; Doyle & Whitely, 1974; Frey, 1973; Kulik & McKeachie, 1975; Leventhal, Abrami & Perry, 1976; Marsh, Fleiner & Thomas, 1975; McKeachie & Kulik, 1975; Sullivan & Skanes, 1974). So, to what conclusion do these validity studies lead? -- Probably to the conclusion that it is important to question the validity of the criterion.

Consider that the construct we want to measure is teaching effectiveness. We have not clearly defined the construct, but we have operationally defined it as student achievement. A common measure of student achievement, at the college level, is grades,

the psychometric properties of which most of the time are unknown. In addition, the relationship between teaching effectiveness and questionably reliable, accurate measures of student achievement has not been empirically tested. To further compound the issue, when we develop a a course evaluation form (CEF), we may or may not include items that go beyond students' knowledge or experience and, therefore may or may not be related to student achievement. For example, students at the introductory level are competent to report what teacher behaviors or course characteristics they observed; they are probably not competent to judge the quality or appropriateness of course content. Although we may call our form a CEF, students may not know: (a) if they are to evaluate the teacher vs. the course or (b) the purpose of their evaluations. Once the CEF is developed, a moderate relationship between students' CEFs and achievement can likely be demonstrated. Voila! Now students' responses to CEFs can be transformed into a number which represents a valid predictor of teaching effectiveness?

To futher complicate the issue, we know there are a variety of factors, which may be situation specific, that have been found to influence the validity of CEF responses. These include student characteristics (e.g., sex, age, grade level), course and class characteristics (e.g., class size, required vs. elective courses, subject matter), teacher characteristics (e.g., sex, faculty rank, personality factors), and logistics of CEF administration (e.g., timing). With the exception of student expectations, student characteristics have little effect on their ratings (Costin, Greenough & Menges, 1975; Leventhal, Abrami,

Perry & Breen, 1975; Perry, Niemi & Jones, 1974). Course and class characteristics have less effect than might be expected. Class size makes a difference if there are 15 or fewer students; required courses are rated lower than elective courses (Centra, 1978; Kulik & McKeachie, 1975). Teacher characteristics show relatively small effects. It may be reasonably concluded that highly rated teachers come in all styles (Kulik & McKeachie). The timing of course evaluations does not appear to be a critical factor (Frey, 1976); however, the purpose of course evaluations does seem to influence ratings. If students know that their ratings will be used to influence personnel decisions, then ratings are higher than otherwise (Aleamoni & Hexner, 1973).

In sum, the most commonly used criterion for demonstrating the validity of CEFs is student achievement. The validity of that criterion for predicting teaching effectiveness has not been tested. Indeed, while we may find that CEFs are differentially valid for various situations, it seems safe to say that two factors particularly influence students' ratings: (a) their knowledge of the purpose for completing the CEF, i.e., the intended audience (Aleamoni & Hexner, 1973) and (b) their expectations (Perry, Niemi & Jones, 1974).

The Evaluation Audiences

As was pointed out above, that evaluation data may have multiple audiences is widely accepted (e.g., House, 1986; McKeachie, 1979; Peterson, Gunne & Rivera, 1984; Scriven, 1974; 1981; 1985). When it comes to the evaluation of teaching effectiveness in higher education, the most commonly described audiences are administrators, instructors, and students. The

different audiences have different needs: (a) administrators need summative information for purposes of making personnel decisions, e.g., promotion, tenure; (b) instructors need formative information for purposes of improving their courses; (c) students need consumer information for purposes of selecting instructors or classes to meet their needs and expectations. Multiple evaluation audiences having different needs and purposes dictate the collection of several different kinds of evaluation data.

Administrators. Administrators need information that will allow them to make summative, yes/no personnel decisions. It has been suggested that the most useful CEF data for administrators is an overall, global rating which describes an instructor's overall teaching effectiveness. Specific items pertaining to the concerns of students (e.g., instructor availability) or instructors (e.g., perceived clarity of presentation) may distract the administrator from the decision at hand because: (a) the validity of any specific item for guaging teaching effectiveness may be questionable, (b) the various audiences may not agree on the importance of individual items as indicators of teaching effectiveness, (c) it is difficult to compare instructors on the basis of differential ratings on various specific items, and (d) incompetent evaluators may weight trivial items more heavily than items representative of overall patterns (Peterson et al., 1984). In considering the global rating, it is important for administrators to consider the knowledge, experience, and expectations that students bring to the evaluation situation. Do students know the purpose(s) of

their evaluations? Are they competent to judge course content vs. teacher behaviors and perceived attitudes? What expectations did the students have of themselves and/or for the course?

Instructors. Instructors usually want evaluative information for purposes of improving their instruction. Gleaning such information requires the use of several specific items which are of importance to the instructor (e.g., perceived availability, clarity of class presentations, impact of exam feedback). As with administrators, it may be important for instructors to consider that students may not be competent to provide them with all the desired sorts of evaluative information. Colleagues may need to be consulted regarding the appropriateness of course content, examinations, etc. McKeachie (1979) further suggests that an important educational goal is to help students develop the ability to evaluate their own learning. This goal can be assessed via CEFs by including items that encourage students to consider what they have learned and their responsibility in the teaching-learning process. Schooling involves learning as well as teaching. What students expected from the course compared to their perceptions of what they got might help instructors put student evaluation data into perspective.

Students. Students are interested in course evaluation data for purposes of selecting classes and instructors. They are the consumers in the evaluation market who, unlike administrators, need information, not conclusions. For example, it may be more useful to students to know that a teacher primarily uses self-paced instruction vs. a lecture format vs. a

discussion format than the overall conclusion that the instructor is a "six" on a seven-point scale. Different students may use the same information to make different decisions that are in line with their expectations.

Can We Move Toward A Plausible Problem Solution?

Peterson et al. (1984) seem to be on a "right track" when it comes to considering the validity of CEF components for various audiences. They have developed a Multiple Audience Rating Form consisting of 17 seven-point likert-type items: (a) one item asks students to rate the instructor's overall teaching effectiveness, (b) nine specific items assess students' concerns and (c) seven items assess teaching performance. The correlations between the global item and each of the other items were positive and significant ($r = .42 - .80$; $n = 1,378$). It was suggested that the results of the three types of items be shared only with the appropriate audience(s). Presumably, administrators would recieve only the global rating, students would receive only the items related to their concerns, and instructors would receive feedback on all of the items because all items function as formative evaluation data for instructors.

When it comes to the validity of using student achievement as the criterion for CEFs designed to measure teaching effectiveness, we are not very far along the road to proven solutions. Essentially, student achievement must be tested as a significant predictor of teaching effectiveness, a construct that remains to be defined. Regardless of the audience, surely the ultimate goal of evaluating teaching effectiveness is to improve the quality of teaching. That common goal is more important than

the technology of developing CEFs. Keeping the goal clearly in mind should facilitate the generation of more than one right problem solution.

Where Do We Go From Here?

A possible approach to the problem of defining teaching effectiveness might be to take a step back and examine from a different perspective what is happening in the classroom. For instance, instead of examining environmental or situational frame factors as previously discussed, a more fruitful tact may be to more closely examine the cognitive processes occurring in the classroom. Doyle (1977) and others (e.g., Morine-Dershimer, 1985; Peterson, Wilkinson & Hallinan, 1984; Stubbs, 1976; Weinstein, 1983; Whitfield, 1976; Winne & Marx, 1983) have considered what students think about during instruction. Particularly, Doyle has argued persuasively for considering the student-centered processes that mediate between instruction and student outcomes, including attitude as well as achievement. Central to his findings is that early in the school experience, students have a variety of thoughts or perceptions of classroom life which influence what and whom they hear in classroom discourse.

The nature of students' thoughts and perceptions which presumably mediate learning outcomes is under study. The mediations seem to be quite diverse. For example: (a) Morine-Dershimer (1985) postulates that students' perceptions of, and participation in, classroom dialogue or discourse mediates success in school; (b) Peterson and Wilkinson (1984) hypothesize that student diversity, teacher-student interaction,

student-student interaction, and classroom organization influence student outcomes; and (c) Doyle (1977) views classroom processes as a two-way (teacher to student and student to teacher), rather than a one way (teacher to student) model. If the current literature on the various operations which characterize classroom interaction has contributed anything, it is the demonstration that: (a) classrooms are interactive environments wherein the teacher and students share responsibility for continuation and maintenance of the environment and (b) that the shared responsibility relates to student outcomes.

The two-way model (Doyle, 1977) of classroom interaction, well-known by those who work in classrooms, but relatively recent to the educational research community, causes one to reconsider the influences on the presentation and achievement of subject matter. How students pass the time spent in class is as important to achievement as is how the teacher spends his/her time. The importance of Doyle's work lies in its construction of the literature on teacher effectiveness. Allocated time is an important correlate of student outcomes; time actually spent on task, both physically and mentally, is of equal importance (Gettinger, 1985). This notion of shared teacher-student responsibility for learning suggests that we cannot validly evaluate instruction unless we know something about the thoughts which students naturally have about instruction.

Therefore, we have a new methodological concern related to assessment of effective teaching -- how might data on mental processes be collected? One method for collecting such data is the stimulated recall interview. Stimulated recall is a

self-report procedure in which a videotape of a class is made and is replayed to each student to stimulate recall of his or her thoughts during the lesson. The videotape is stopped periodically during the playback and the student is interviewed by the researcher in an attempt to capture those thoughts. Thus far, in this type of research, stimulated recall has been generally accepted as an effective method. It yields information about the participants' thoughts during the class period, their reporting what is heard, or what happened, and of course, what is not heard, or did not happen.

The available data demonstrate: (a) a wide variety of thoughts occur during instruction that are developed early in the school experience and (b) that students continue to create, not so much a set of expectations, as they do a set of filters. The following studies are offered as exemplars of the effective use of the stimulated recall interview. While it cannot be assumed that the results of these representative studies are generalizable to higher education classrooms, it does appear that stimulated recall may prove useful in determining why students evaluate a course in a particular way.

Research Using Stimulated Recall

Morine-Dershimer (1985) studied pupil perceptions of classroom discourse. She found that the discourse related to the material under study was perceived by elementary school students as separate from their definition of the "important talk" in the classroom. Further, regardless of students' degree of mental participation during instruction, participation in classroom discussions was found to be related to academic achievement.

According to Mergendoller and Packer (1985) many of the studies of student perceptions do not validly reflect such perceptions because students are forced to respond in ways that fit the needs of the researcher. To overcome this problem, they interviewed 20 seventh grade students to determine what words the students frequently used to describe their teachers. In short, eight frequently used words were: mean, hard, easy, good, strict, boring, fun, and nice. Follow-up interviews revealed that the frequently used words represented four basic themes: academic work, instructional facility, classroom experience, and personal characteristics. Girls mentioned instructional facility more frequently than did boys; boys mentioned academic work more often. Like Morine-Dershimer (1985), Mergendoller and Packer concluded that students form conceptions of teachers early. They were additionally struck by the clarity and frankness with which the students could explain themselves and their evaluations and feelings about the teacher. A question worthy of future study is whether students' conceptions affect their subsequent performance.

In developing the Adjective Rating Scale (ARS), Kelly, Pascarella, Terenzini, and Chapman (1976) used research methods similar to those used by Mergendoller and Packer (1985). University students generated adjectives for reporting their attitudes toward a course. It was assumed that frequently used words were representative of mental constructs. The resultant ARS consisted of 24 adjectives that fell into four factors; interest value, dullness, practical appeal, and difficulty. Evidence provided by Kelly et al. indicates that

students' responses on the ARS were modestly related to academic achievement as measured by freshman grade point average (GPA) and amount of nonformal contact with faculty.

The studies described above illustrate that students develop a set of mental mediators early in their school experience. These mediators or filters develop the skill of selective attention. The students thereby know what is important to them and how to describe the processes of the classroom.

It is hypothesized, then, that these mediators are also at work when students are asked to evaluate teaching. Frey's (1976) finding that the timing of an evaluation is of little consequence is encouraging when considered in light of this perspective on the research. That evaluations given anytime after the sixth week and before the thirteenth week yield similar data supports the position taken here that students form conceptions early and that they are stable. Faculty evaluation schemes that fail to consider the students' cognitions, or the students' rationale for making a particular judgment are neglecting valuable information about instruction. What do students think about when they respond to the item, "This teacher is effective at providing examples" (or any other evaluation statement)? How do they define "effective"? It is a concept; it has examples and attributes. What are they? It would seem that such data are equally as important as the rating itself. Useful factors such as those identified by Mergendoller and Packer (1985) or Kelly, et al. (1986) will continue to populate the literature, however, they are only of moderate help in understanding faculty evaluation. If the practice of faculty evaluation is to continue

to use students as informants, then a commitment should be made to extract as much characteristic data from the students as possible to provide the data that we typically do not have. Understanding how students mediate the purpose and content of instruction seems to be one promising way to uncover some other explanations of faculty evaluations.

Where do we go from here? Perhaps we begin to look at the evaluation of teaching effectiveness from a different perspective. We have long recognized the influence of environmental and personological frame factors on student products, including students' responses to CEFs. Given the current interest in, and developing technology of, cognitive science, we may now be ready to start exploring the impact of the interaction between cognitive variables and identified frame factors on students' evaluation products. Consideration of cognitive variables, as they likely interact with other variables, may enlighten our attempts to define and, hence, validly measure teaching effectiveness.

References

Aleamoni, L.M., & Hexner, P.Z. (1973). The effect of different sets of instructions on student course and instructor evaluation, Research Report No. 339, Urbana, IL: Measurement and Research Division, Office of Instructional Resources, University of Illinois.

Braskamp, L.A., Brandenburg, D.C., & Ory, J.C. (1984). Evaluating teaching effectiveness: A practical guide. Beverly Hills, CA: Sage

Centra, J.A. (1977). Student ratings of instruction and their relationship to student learning. American Educational Research Journal, 14, 17-24.

Centra, J.A. (1978). Using student assessments to improve performance and validity. New Directions for Institutional Research, 20, 31-49.

Centra, J.A. (1983). Research productivity and teaching effectiveness. Research in Higher Education, 18, 379-389.

Costin, F., Greenough, W.T., & Menges, R.J. (1971). Student ratings of college teaching: Reliability, validity, and usefulness. Review of Educational Research, 41, 511-535.

Doyle, K.O. (1975). Student evaluation of instruction,

Lexington, MA: Lexington Books.

Doyle, W. (1977). Paradigms for research on teacher effectiveness. In L.S. Shulman (Ed.), Review of Research in Education. Itasca, IL: Peacock.

Doyle, K.O., & Whitely, S.E. (1974). Student ratings as criteria for effective teaching. American Educational Research Journal, 11, 259-274.

Feldman, K.A. (1983). Seniority and experience of college teachers as related to evaluations they receive from students. Research in Higher Education, 18, 3-124

Frey, P.W. (1973). Student ratings of teaching: Validity of several rating factors. Science, 182, 83-85.

Frey, P.W. (1976). Validity of student instructional ratings: Does timing matter? Journal of Higher Education, 47, 327-336.

House, E.R. (1986). In-House reflections: Drawing evaluative conclusions. Evaluation Practice, 7(3), 35-39

Kelly, E.F., Pascarella, E.T., Terenzini, P.T., & Chapman, D. (1976). The development and use of the Adjective Rating Scale: A measure of attitude toward courses and programs. Syracuse, NY: Center for Instructional Development.

Kinicki, A. J., Bannister, B.D., Hom, P.U., Denisi, D.G. (1985). Behaviorally anchored rating scales vs summated rating scales: Psychometric properties and susceptibility to rating bias. Educational and Psychological Measurement, 45, 535-549.

Kulik, J.A., & McKeachie, W.J. (1975). The evaluation of teachers in higher education. In F.N. Kerlinger (Ed.), Review of research in education, volume 3, Itasca, IL: Peacock.

Leventhal, L., Abrami, P.C., & Perry, R.P. (1976). Do teacher rating forms reveal as much about students as about teachers? Journal of Educational Psychology, 68, 441-445.

Leventhal, L., Abrami, P.C., Perry, R.P., & Breen, L.J. (1975). Section selection in multi-section courses: Implications for the validation and use of teacher rating forms. Educational and Psychological Measurement, 35, 885-895.

Lin, Y., McKeachie, W.J., & Tucker, D.G. (1984). The use of student ratings in promotion decisions. Journal of Higher Education, 55, 583-589.

Marsh, H.W., Fleiner, H., & Thomas, C.S. (1975). Validity and usefulness of student evaluations of instructional quality. Journal of Educational Psychology, 67, 833-839.

Marsh, H.W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential baises, and utility. Journal of Educational Psychology, 76, 707-754.

McCaleb, J. & White, J. (1980). Critical dimensions of evaluating teacher clarity. Journal of Classroom Interaction, 15, 27-30.

McCarthy, P.R., & Schmeck, R.R. (1982). Effects of teacher self-disclosure on student learning and perceptions of teacher. College Student Journal, 16, 45-49.

McKeachie, W.J. (1979). Student ratings of faculty: A reprise. Academe, October 1979, 384-397.

McKeachie, W.J., & Kulik, J.A. (1975). The effectiveness of instruction in higher education. In F.N. Kerlinger (Ed.), Review of research in education, volume 3, Itasca, IL: Peacock.

Mergendoller, J.R., & Packer, M.J. (1985). Seventh graders' conceptions of teachers: An interpretive analysis. Elementary School Journal, 85(5), 581-600.

Miller, S. (1984). Student rating scales for tenure and promotion. Improving College and University Teaching, 32, 87-90.

Morine-Dershimer, G. (1985). Talking, listening, and learning in elementary classrooms. New York: Longman.

Murray, H.G. (1983). Low-inference classroom teaching behaviors and student ratings of college teaching effectiveness. Journal of Educational Psychology, 75, 138-149.

Perry, R.P., Niemi, R.R., & Jones, K. (1974). Effect of prior teaching evaluations and lecture presentation on ratings of teaching performance. Journal of Educational Psychology, 66, 851-856.

Peterson, K., Gunne, G.M., Miller, P., & Rivera, O. (1984). Multiple Audience Rating Form strategies for student evaluation of college teaching. Research in Higher Education, 20 (3), 309-321.

Peterson, P.L., Wilkinson, L.C., & Hallinen, M. (1984). The social context of instruction. Orlando, FL: Academic Press.

Scriven, M. (1974). The evaluation of teachers and teaching. California Journal of Educational Research, 24 (3), 109-118.

Scriven, M. (1981). Summative teacher evaluation. In J. Millman (Ed.), Handbook of teacher evaluation. Beverly Hills, CA: Sage Publications.

Scriven, M. (1985). New frontiers of evaluation. Evaluation Practice, 7 (1), 7-44.

Stubbs, M. (1976). Language, schools, and classrooms. London: Methuen.

Sullivan, A.M., & Skanes, G.R. (1974). Validity of student evaluation of teaching and the characteristics of successful instructors. Journal of Educational Psychology, 66, 584-590.

Weinstein, R. (1983). Student perceptions of schooling. Elementary School Journal, 83 (4), 287-312.

Whitfield, T. (1976). How students perceive their teachers. Theory into Practice, 15, 347-351.

Winne, P.H., & Marx, R.W. (1982). Students' and teachers' views of thinking processes involved in classroom learning. Elementary School Journal, 82 (5), 493-518.